Using retinex for point selection in 3D shape registration

Liu, Yonghuai; Martin, Ralph R.; de Dominicis, Luigi; Li, Baihua

Published in:
Pattern Recognition

DOI:
10.1016/j.patcog.2013.12.015

Publication date:
2014

Citation for published version (APA):
Liu, Y., Martin, R. R., de Dominicis, L., & Li, B. (2014). Using retinex for point selection in 3D shape registration. *Pattern Recognition*, *47*(6), 2126-2142. https://doi.org/10.1016/j.patcog.2013.12.015

Using Retinex for Point Selection in 3D Shape Registration

Yonghuai Liu[a], Ralph R. Martin[b], Luigi de Dominicis[c], Baihua Li[d]
[a]Department of Computer Science
Aberystwyth University, Ceredigion SY23 3DB, UK
Email: yyl@aber.ac.uk
[b]School of Computer Science & Informatics
Cardiff University, Cardiff CF24 3AA, UK
Email: ralph@cs.cf.ac.uk
[c]Diagnostics and Metrology Laboratory UTAPRAD-DIM
00044 ENEA Frascati, Italy
Email: luigi.dedominicis@enea.it
[d] School of Computing, Mathematics & Digital Technology
Manchester Metropolitan University, Manchester M1 5GD, UK
Email: b.li@mmu.ac.uk

Abstract

Inspired by retinex theory, we propose a novel method for selecting key points from a depth map of a 3D freeform shape; we also use these key points as a basis for shape registration. To find key points, first, depths are transformed using the Hotelling method and normalized to reduce their dependence on a particular viewpoint. Adaptive smoothing is then applied using weights which decrease with spatial gradient and local inhomogeneity; this preserves local features such as edges and corners while ensuring smoothed depths are not reduced. Key points are those with locally maximal depths, faithfully capturing shape. We show how such key points can be used in an efficient registration process, using two state-of-the-art iterative closest point variants. A comparative study with leading alternatives, using real range images, shows that our approach provides informative, expressive, and repeatable points leading to the most accurate registration results.

Keywords: Retinex; Key point; Freeform shape; Adaptive smoothing; Registration

1 Introduction

Laser scanning systems can readily capture the surfaces of 3D shapes within minutes, with results like those in Figure 1. Since scanners have a limited field of view, and one part of the shape may occlude others, multiple datasets must be captured from different viewpoints to obtain (approximately) full coverage of the shape of interest. These datasets are in the form of *depth maps* in local scanner-centric coordinates. In order to fuse the information in these datasets to give a single complete surface, they must be brought into a single global coordinate system using a process of *registration*. This aims to find pairwise underlying transformations that align one dataset with another. Typically, prior knowledge of the underlying transformation is lacking, and furthermore, occlusion causes appearance and disappearance of points in different datasets. Registration is thus a challenging problem which has attracted attention in various disciplines, such as computer vision, pattern recognition, computer graphics, and

medical imaging. It finds applications in such areas as object recognition, quality assurance, computer aided design, medical diagnosis, and therapy planning. Many algorithms have been developed for the registration of overlapping 3D freeform shapes [5, 9, 18]. In this paper, our main concern is the detection of *key points* for *faithful representation* and *efficient registration* of 3D freeform shapes.

1.1 Previous work

The detection of key points which capture the important features of a freeform shape is of great interest, since they provide a compact alternative to the original geometry, and may be used in its place in various tasks such as shape registration, data transmission, rendering and visualisation. Various methods have been proposed for the detection of key points. They should capture both significant local details of the shape and its overall geometry. Their determination should be repeatable and they should represent intrinsic features: key points detected from one viewpoint should ideally be the same as those detected from another. Key points should also be resistant to imaging noise and resolution, and robust in the presence of occlusion of parts of the object. Finally, their detection should be computationally efficient.

Normal space sampling (NSS) was proposed in [20] as a way of sampling informative points. It calculates angles between the normal vectors at a point of interest and its neighbors, and uses these angles as a basis for sampling. More points are sampled in areas where normal vectors change more rapidly. A relevance based sampling scheme was proposed in [28], which defines the distinctiveness of a point in terms of a surrounding area having similar normal vectors. The larger the area, the less distinct the point. To address the issue of different densities of point sampling due to variations in the distance between the object and the scanner, a spherical subsampling method was proposed in [12]: instead of sampling points from regularly gridded data, it considers the scanner's angular resolution. In [13], a Hotelling transform is first performed over the points; feature points are detected as those with the largest difference between the ranges spanned by the x and y coordinates in the transform. A supervised learning method was proposed in [3] for feature point localization within human faces. Gaussian mixture models (GMMs) are used to approximate the distribution of candidate feature points; a 3D model of feature points is built which enforces constraints on localization of feature points. Another learning method was proposed in [4] for detecting key points in faces. A face is represented using 14 landmarks, and statistical distributions of their descriptors and weights used for linear combination are learnt from training data. For a given face mesh, a number of descriptor maps are computed and matched against the 14 learnt ones, and combined using the learnt weights. The key points are determined as strong local maxima above a threshold of the combined values over the 14 landmark dictionary shapes. The classical Harris corner detector for 2D images was adapted in [22] to 3D meshes. Principal component analysis is first performed over local patches, then all points on a patch are transformed so that its normal vector is a local z axis, leading the points to lie close to the xy plane. A paraboloid is fitted to these transformed points, allowing x and y derivatives to be computed. To address the difference between discrete images and continuous patches, derivatives are convolved with a Gaussian function. An autocorrelation matrix is used to estimate the local corner strength. A technique called accumulative geodesic extrema (AGEX) was proposed in [19] for human body feature point detection. Key points are those with the longest distances on the surface mesh from its geodesic centroid, distances being computed using shortest geodesic paths.

The gram matrix was analysed in [14] as a means of geometric corner detection. A set of corner points at each scale is found by searching for local spatial maxima of corner detector responses. Corners lying along edge points are pruned by thresholding the variance of the second-order partial derivatives. The octave method of salient point detection was proposed in [2], based on the projection of the difference

of the weighted averages of the points inside spheres of varying radii onto the weighted average of the normal vectors of the points inside a sphere with smaller radius. A multiscale feature extraction method was proposed in [17] based on principal component analysis, using the ratio of the resulting eigenvalues as an indicator of saliency. Multiscale analysis is also performed in [11] for key point detection in a range image. It first segments the range image to isolate each area of interest, then depth is normalized to a global average of zero and a standard deviation of 1. Multiscale Gaussian and difference of Gaussian (DoG) analysis is performed over these normalized depth values. A key point is one at a local extremum in the DoG space, larger than a threshold. Key points are detected in [26] from regions with significant variation in mean curvature. It first constructs a structure tensor in each local region as a function of the local mean curvatures and normal vectors; it is invariant to the sampling density. It then detects corners at the spatial extrema of the determinant of the structure tensor, and finally rejects spurious key points with low response, an edge response, or near to a depth discontinuity. A variant of the Laplacian of Gaussian method was proposed in [25] for key point detection. It first calculates the depth gradient along both x and y directions, then the depth gradient is normalized taking into account the angular resolution of the range image, and the second order derivative is estimated from the normalized depth gradient, yielding the magnitude of the gradient in the unit interval. A high magnitude marks a possible key point with high curvature, while spurious points are filtered by occlusion analysis and linear structural analysis. Another salient point detector was proposed by the same author in [24] which first uses principal component analysis to estimate the local surface variation at a particular point and its main direction. Key point strength is calculated by considering how much the main directions change from each other and how stable this point is on the surface. These values are smoothed over the whole image, and points with local maxima larger than a threshold are selected as key points. In [23], given an intensity image associated with a range map, the image mesh is generated and smoothed using a multiple scale bilateral filter, then the gradient at each vertex is estimated using the Laplace-Beltrami operator (LBO). Points with locally extremal gradients are filtered by thresholding their LBO response and suppressed by a non-maximal scheme to finally detect key points.

The above methods can be classified into two main categories, depending on whether they use single scale or multi-scale analysis. While the former are usually more computationally efficient, the latter are more robust to imaging noise, changes in resolution, and occlusion. They often involve the computation of second order derivatives of shape, filtering spurious salient points, and sometimes, learning. Unfortunately, second order derivatives are sensitive to occlusion, depth discontinuities, changes in sampling resolution and imaging noise. Thresholds are difficult to select and data dependent, while learning requires many repeatable samples which may be difficult or impossible to collect. The detection of key points from 3D freeform shapes is challenging and still remains open.

If salient key points can be reliably found, subsequent range image analysis may be simpler and more efficient. Conversely, such range image analysis results may provide an effective tool to quantitatively evaluate the extent to which useful key points have been reliably detected.

1.2 Our work

Retinex theory [6, 16] considers how brightness and reflectance behave, and investigates a computational model of color constancy: human perception of colors is largely independent of illumination conditions. It shows that a captured 2D image can be decomposed into two subimages: one depends on the reflectance properties of the surface of the imaged object, while the other depends on the illumination conditions. If such a decomposition can be computed in practice, the reflectance image can be used to improve reliability of such tasks as face recognition, as this image more faithfully represents the geo-

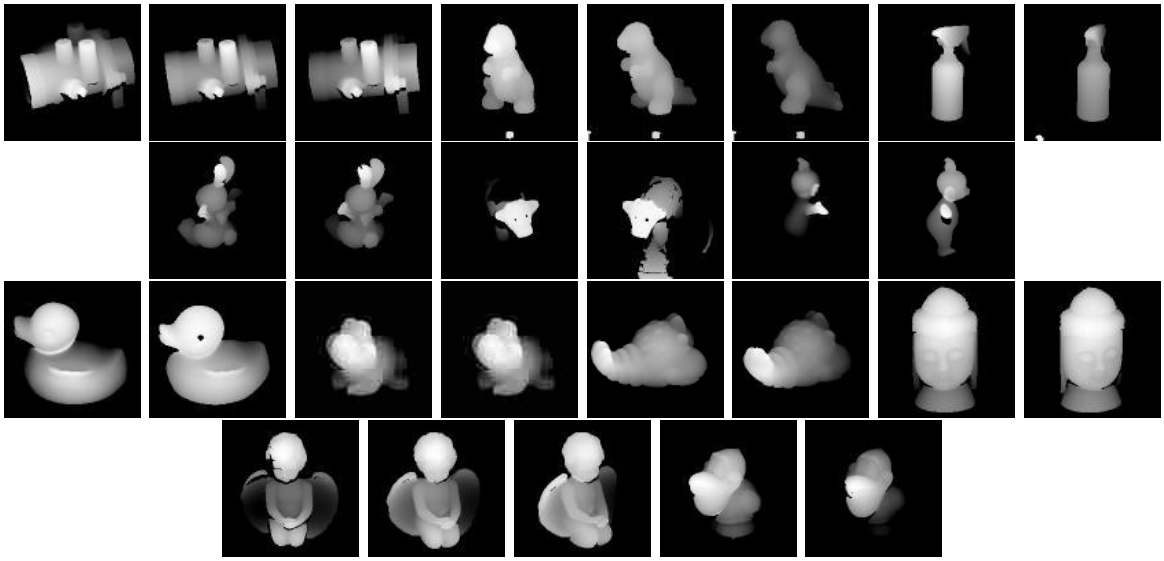

Figure 1: Real range images used. Top: valve20, valve10, valve0, dinosaur72, dinosaur36, dinosaur0, bottle0, and bottle36; Second: bunny80, bunn60, cow49, cow45, tubby120, and tubby80. Third: duck0, duck20, frog0, frog20, lobster0, lobster20, buddha0, and buddha20. Bottom: angel0, angle20, angel40, bird0, and bird20.

metrical properties of the imaged object, while the variable effects of illumination have been removed. While such image decomposition is an ill-posed problem, various approaches have been proposed based on Gaussian smoothing [6], adaptive smoothing [16] and minimization of the sum of the first order derivatives of the illumination and the difference between the illumination and the given image [21, 8]. Two ideas are of particular interest: adaptive smoothing [16] and reflectance inequality [21, 8], where the reflectance component R is estimated as the logarithm of the ratio of intensity F of the pixel of interest and that L of its neighbors as illumination component, satisfying the constraint $R \leq 1$ and thus, $F \leq L$. Using these ideas, we propose a novel method in this paper for the detection of key points on a 3D freeform shape. To this end, depth values are first transformed and normalized so that their dependence on the particular viewpoint can be reduced. Adaptive smoothing is then applied to the normalized depths with weights defined as decreasing functions of spatial gradient and local inhomogeneity. This adaptive smoothing makes sure that local features such as corners and edges are preserved, while smoothed depths become insensitive to imaging noise, but are not reduced. Finally, key points are detected as those unaffected by the adaptive smoothing process and thus have locally maximum transformed and normalized depths.

These detected key points are intended for use as proxies to represent the original shape; here, we investigate whether they can represent the original shape faithfully and are suitable for solving the registration problem. If they can, then the computational efficiency of registration of overlapping 3D depth maps can be significantly improved. We carry out a comparative study, using three other state-of-the-art salient point selection methods: the octave algorithm [2], a multi-scale feature extraction (MSFE) method [17], and the normal space sampling (NSS) method [20]. The octave and MSFE methods are multi-scale methods, while NSS is a single scale method. This comparative study reveals which method can best detect informative, expressive and repeatable feature points. To determine the utility of the detected key points for downstream applications, two state-of-the-art iterative closest point (ICP) variants, SoftICP [9] and fractional RMSD (FICP) [18], were used for registration. The former

is an extension of the SoftAssign algorithm [5] which applies the entropy maximization principle to estimate weights of different tentative correspondences, then uses a two-way constraint to refine these weights before estimating the underlying transformation. To ensure robust results, these two steps are embedded in a deterministic annealing scheme. The FICP algorithm simultaneously optimizes both the size of the overlap between depth images, and the transformation parameters.

To assess the performance of the key point detection algorithms, the following measurements were made: precision and recall rates of detected key points, average e_μ and standard deviation e_σ of registration errors in millimetres for reciprocal correspondences (RCs) [9, 27], expected and estimated rotation angles θ and $\hat{\theta}$ in degrees of the underlying transformation, and the time taken for automatic key point detection and registration. While the precision and recall rates measure the repeatability of the detected key points [22, 24, 26], the corresponding registration results measure the informativeness and expressiveness of these key points. As the same registration algorithms were used in each case, differences in their performance come solely from the types of points used. The better the registration results obtained, the more representative we can consider the selected points to be of the shapes.

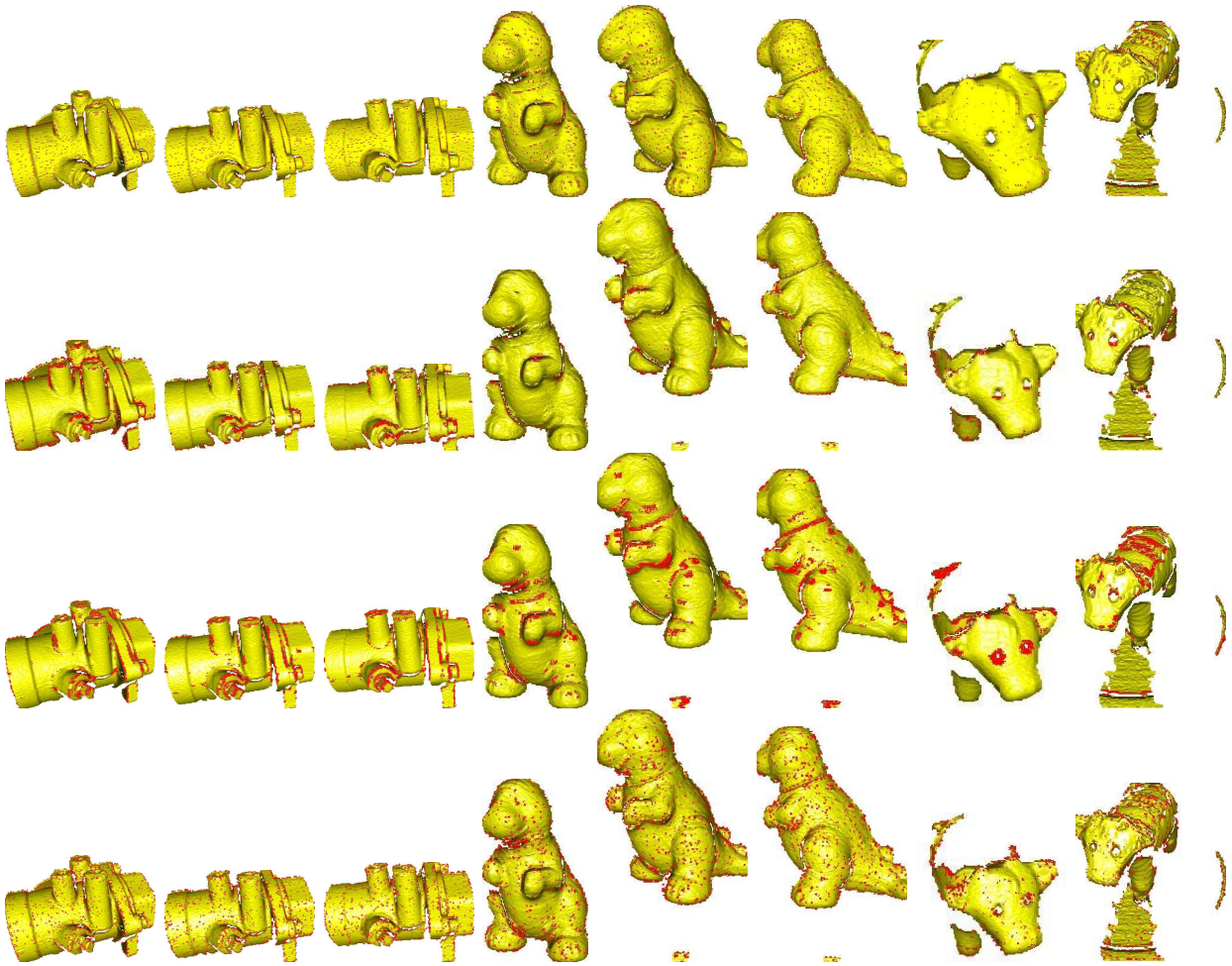

Figure 2: Key points detected in different images using different algorithms. Columns, left to right: valve20, valve10, valve0, dinosaur72, dinosaur36, dinosaur0, cow49, and cow45. Rows, top to bottom: results using our method (RKP), octave, MSFE, and NSS.

In the remainder of the paper, Section 2 describes our key point detection method, Section 3 presents experimental results, and Section 4 draws some conclusions and indicates future work.

2 A novel method for 3D key point detection

The following notation is used throughout this paper: bold face letters denote vectors, matrices, or sets, $|\cdot|$ denotes the absolute value of a scalar or the number of elements in a set, $\mathbf{a} \cdot \mathbf{b}$ denotes the dot product of vectors $\mathbf{a}$ and $\mathbf{b}$, $||\cdot||$ denotes the Euclidean norm of a vector, and superscript ^{T} denotes the transpose of a vector.

A freeform shape as illustrated in Figure 1 is represented by four arrays, each of size of $H \times W$: flag f, x coordinate, y coordinate, and z coordinate. If $f(i,j) = 1$, then the point at location (i,j) is valid and can be extracted as $(x(i,j)\ y(i,j)\ z(i,j))^T$, otherwise, it should be ignored. An invalid point is one for which no reliable coordinates on the imaged object surface can be estimated by the scanner (because the reflected signal received is either too weak or too strong due to low reflectance or specular reflection).

We now explain our key point detection method, concentrating on four main issues: depth normalization, weight estimation, adaptive smoothing, and key point extraction.

2.1 Depth normalization

Even though the depth information $z(i,j)$ directly represents the given shape, it is viewpoint dependent. To facilitate key point detection, the Hotelling transform [13] is applied to reduce such viewpoint dependence. Thus, for each valid pixel (i,j) inside the image of the given shape, all valid neighboring points inside a window of size of $s \times s$ are extracted:

$$\mathbf{A} = \{\mathbf{a}_i\} = \{(x(i+k,j+l)\ y(i+k,j+l)\ z(i+k,j+l)^T \mid f(i+k,j+l) = 1, -s/2 \le k,l \le s/2\}.$$

Let the centroid of all the points in $\mathbf{A}$ be $\bar{\mathbf{a}} = \sum_{\mathbf{a}_i \in \mathbf{A}} \mathbf{a}_i / |\mathbf{A}|$. Then their covariance matrix $\mathbf{C}$ can be constructed as: $\mathbf{C} = \sum_{\mathbf{a}_i \in \mathbf{A}} (\mathbf{a}_i - \bar{\mathbf{a}})(\mathbf{a}_i - \bar{\mathbf{a}})^T$. The matrix $\mathbf{C}$ is symmetric; its eigenvectors $\mathbf{v}_i$ and corresponding eigenvalues λ_i can be found using the Jacobi method. Suppose that the $\mathbf{v}_i$ are sorted in descending order of λ_i and are put together into a matrix $\mathbf{v}$ as $\mathbf{v} = (\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3)$. Then the point $\mathbf{p} = (x(i,j)\ y(i,j)\ z(i,j))^T$ at pixel (i,j) can be transformed to: $\tilde{\mathbf{p}} = (\tilde{x}(i,j)\ \tilde{y}(i,j)\ \tilde{z}(i,j))^T = \mathbf{v} \cdot (\mathbf{p} - \bar{\mathbf{a}})$. Next, $\tilde{z}(i,j)$ is normalized over the whole image, setting

$$\tilde{z}(i,j) = 255(\tilde{z}(i,j) - \tilde{z}_{\min})/(\tilde{z}_{\max} - \tilde{z}_{\min})$$

where $\tilde{z}_{\max}$ and $\tilde{z}_{\min}$ are the maximum and minimum values over the whole image. The resulting $\tilde{z}(i,j)$ is used as a normalized viewpoint independent depth for subsequent key point detection. The difference between $\mathbf{p}$ and $\bar{\mathbf{a}}$ removes the effect of translation, while the dot product between $\mathbf{v}$ and $\mathbf{p} - \bar{\mathbf{a}}$ removes the effect of rotation. However, computation of $\bar{\mathbf{a}}$ and $\mathbf{v}$ may be affected by imaging noise, resolution, and occlusion and as a result, the dependence of $\tilde{z}$ on a particular viewpoint is only partially reduced, not completely.

2.2 Weight definition

Key point detection applies an adaptive smoothing operation to the transformed and normalized depth values. This operation should preserve local features such as corners and edges. To do so, the weights for

Figure 3: Registration results for different shapes using all points and different algorithms. Columns, left to right: valve20-10, valve10-0, dinosaur72-36, dinosaur36-0, bottle0-36, bunny80-60, cow49-45, and tubby120-80. Rows, top to bottom: SoftICP, FICP.

smoothing must be appropriately defined, taking into account the spatial gradient and inhomogeneity: the larger these two factors are, the smaller the weight should be. The function $w(r) = 1/(1+\sqrt{r})$, $r \geq 0$ is used to suppress large spatial gradients and inhomogeneity; it has the following useful properties: (i) it is continuous, so does not abruptly change given a small change in r, and (ii) it decreases with r as required.

The local spatial gradient $g(i,j)$ at pixel (i,j) is the square root of the sum of the squares of the differences in depths of pixels in horizontal and vertical directions (thus using the L_2 norm):

$$g(i,j) = \sqrt{g_u^2(i,j) + g_v^2(i,j)}$$

where $g_u(i,j) = \tilde{z}(i+1,j) - \tilde{z}(i-1,j)$ and $g_v(i,j) = \tilde{z}(i,j+1) - \tilde{z}(i,j-1)$. Suppose the average gradient of the whole image is $\bar{g}$. Then the weight $w_g(i,j)$ for the gradient component is set to $w_g(i,j) = 1/(1+\sqrt{0.1g(i,j)\exp(0.1\bar{g})})$.

Given a valid pixel at location (i,j), its valid neighbors are in: $\mathbf{N} = \{(i+k,j+l)|f(i+k,j+l) = 1, -s/2 \leq k,l \leq s/2\}$. The local inhomogeneity $h(i,j)$ at pixel (i,j) considers the difference of depth between neighboring pixels and is set to the average of the absolute values of these differences (using the L_1 norm):

$$h(i,j) = \frac{1}{|\mathbf{N}|} \sum_{(m,n)\in\mathbf{N}} |\tilde{z}(m,n) - \tilde{z}(i,j)|.$$

Then $h(i,j)$ is normalized using: $h(i,j) \leftarrow (h(i,j) - h_{\min})/(h_{\max} - h_{\min})$ where $h_{\min}$ and $h_{\max}$ are the minimum and maximum inhomogeneity over the whole image. To further suppress large inhomogeneities, $h(i,j)$ is transformed to: $\hat{h}(i,j) = \sin(0.5h(i,j)\pi)$. Finally, the weight $w_h(i,j)$ for the inhomogenity component is set to: $w_h(i,j) = 1/(1+\sqrt{10\hat{h}(i,j)\exp(10\bar{h})})$ where $\bar{h}$ is the average of

$\hat{h}(i,j)$ over the whole image.

While the spatial gradient in the L_2 norm is likely to be dominated by noisy pixels or pixels with large depth discontinuities, the spatial inhomogeneity in the L_1 norm is useful to capture small changes in depths of pixels along edges or at corners. Combining the weights $w_g(i,j)$ and $w_h(i,j)$ for the spatial gradient and inhomogeneity components together thus captures both coarse and fine features in the underlying shape, giving the final weight $w(i,j)$ for each valid pixel (i,j): $w(i,j) = w_g(i,j)w_h(i,j)$.

2.3 Adaptive smoothing

Once the weights $w(i,j)$ have been defined, adaptive smoothing is performed. Following the idea used to estimate the illumination component of an image in retinex theory [16, 21], we assume that the smoothed depth cannot be smaller than the original depth. Thus, adaptive smoothing is performed using the following steps:

Initialize the smoothed depth $s_0(i,j) = \tilde{z}(i,j)$, the number of iterations $I = 0$, and the maximum number of iterations $I_{\max}$

While $I < I_{\max}$, do:

$I \leftarrow I + 1$

If the pixel at location (i,j) is valid, extract all its valid neighbors:
$\mathbf{N} = \{(i+k, j+l) \,|\, \mathrm{f}(i+k, j+l) = 1, -s/2 \le k, l \le s/2\}$.
Compute the weighted average $\hat{s}(i,j)$ of smoothed depths $s_0(m,n)$ at pixels (m,n) in $\mathbf{N}$ using weights $w(m,n)$: $\hat{s}(i,j) = \sum_{(m,n)\in\mathbf{N}} s_0(m,n)w(m,n) / \sum_{(m,n)\in\mathbf{N}} w(m,n)$
Ensure the smoothed depths are non-decreasing: $s_1(i,j) = \max(\hat{s}(i,j), s_0(i,j))$

Prepare for the next round: $s_0(i,j) \leftarrow s_1(i,j)$

The parameter $I_{\max}$ affects the number and location of keypoints to be selected; unless otherwise stated, we set $I_{\max} = 30$.

2.4 Key point detection

Pulling all the ingredients defined in the previous sections, the novel key point detection algorithm is summarized as follows:

Initialize window size to $s \times s$; unless otherwise stated, $s = 3$ in this paper

Normalize the depth values of the given range image

Estimate the weight of each valid pixel

Perform adaptive smoothing

Transform the smoothed depths $s_1(i,j)$ to $\tilde{b}(i,j)$ using a logarithmic operation:
$\tilde{b}(i,j) = \log((\tilde{z}(i,j)+1)/(s_1(i,j)+1))$.

Normalize the logarithmically transformed depths:
$\bar{b}(i,j) = (\tilde{b}(i,j) - \tilde{b}_{\min})/(\tilde{b}_{\max} - \tilde{b}_{\min})$
where $\tilde{b}_{\min}$ and $\tilde{b}_{\max}$ are the minimum and maximum of $\tilde{b}(i,j)$ over the whole image.

Table 1: Precision Pre and recall Rec rates of the detected keypoints, the average e_μ and standard deviation e_σ of registration errors in millimetres based on RCs, expected and estimated rotation angles θ and $\hat{\theta}$ in degrees, and registration time t in seconds for different algorithms and different freeform shapes.

Shapes	Algo.	Pre (%)	Rec (%)	e_μ (mm)	e_σ (mm)	θ $(^\circ)$	$\hat{\theta}$ $(^\circ)$	t (s)
valve20-10	SoftICP	76.70	84.88	0.40	0.22	10	10.10	97
	FICP	76.95	85.16	0.40	0.22		10.12	77
valve10-0	SoftICP	86.16	86.75	0.38	0.21	10	10.11	66
	FICP	86.56	87.16	0.39	0.21		10.11	45
dinosaur72-36	SoftICP	60.52	51.51	0.62	0.87	36	35.15	41
	FICP	61.00	51.92	0.63	0.85		35.02	34
dinosaur36-0	SoftICP	63.35	61.02	0.56	0.54	36	35.72	66
	FICP	63.58	61.24	0.56	0.54		35.66	50
botttle0-36	SoftICP	59.08	62.14	0.67	0.33	36	32.68	31
	FICP	70.91	74.59	0.67	0.58		3.88	30
bunny80-60	SoftICP	76.76	71.87	0.22	0.11	20	19.94	29
	FICP	76.49	71.61	0.22	0.11		19.87	20
cow49-45	SoftICP	47.81	21.36	0.71	0.70	40	41.30	26
	FICP	46.95	20.98	0.71	0.71		41.59	28
tubby120-80	SoftICP	50.60	44.42	0.25	0.16	40	39.09	23
	FICP	50.57	44.39	0.26	0.17		39.02	30

Extract key points as those $(x(i,j)\ y(i,j)\ z(i,j))^T$ for which $\bar{b}(i,j) = 1$.

We call the above algorithm the *retinex key point* (RKP) algorithm. It has a computational complexity of $O(n)$ for depth transformation, normalization, and weight estimation, $O(I_{\max} n) = O(n)$ for adaptive smoothing, and $O(n)$ for key point extraction. Overall, it thus has linear computational complexity $O(n)$ in the number n of valid points in the shape. This algorithm has the following property:

Property 1 *Detected key points have locally maximal transformed and normalized depths.*

Proof: The definition of a key point as one satisfying $\bar{b}(i,j) = 1$ means that $\tilde{b}(i,j) = \tilde{b}_{\max}$. The non-decreasing constraint on depth during adaptive smoothing implies that $\tilde{z}(i,j) \leq s_1(i,j)$ and thus $\tilde{b}(i,j) \leq 0$. Consequently, $\tilde{b}_{\max} = 0$ and also $\tilde{b}(i,j) = 0$, so $\tilde{z}(i,j) = s_1(i,j)$. From the initial values of $s_0(i,j) = \tilde{z}(i,j)$ and the non-decreasing constraint on smoothed depth, $s_1(i,j) = \max(\hat{s}(i,j), s_0(i,j))$, we have $\tilde{z}(i,j) = \max(\hat{s}(i,j), \tilde{z}(i,j))$. Thus, pixel (i,j) has a locally maximal transformed and normalized depth, a property which is unaffected by adaptive smoothing. The adaptively smoothed depth determines whether a point has a locally maximal depth. ■

3 Experimental results

In this section, we use real data to demonstrate the utility of our algorithm for key point detection and its application for efficient registration of overlapping 3D freeform shapes. The detected key points are

directly used for registration in two state-of-the-art iterative closest point (ICP) variants: SoftICP [9] and Fractional RMSD (FICP) [18]. The RKP algorithm usually selects around 10% points in the shape, so to enable a fair comparison in registration, we have also ensured that the octave, MSFE and NSS methods all select 10% of the points from the shape.

For each pair of overlapping shapes, we refer to the first as the *data shape*, and the second as the *reference shape*. All real data in Figure 1 were downloaded from [15]. They were captured using a Minolta Vivid 700 range camera with a resolution of 200×200. The performance of the algorithm is measured using the following parameters: the precision Pre and recall Rec rates as a percentage of the detected keypoints, the average and standard deviation of registration errors of reciprocal correspondences (RCs) [9, 27] between the whole data and reference shapes, the rotation angle $\hat{\theta}$ in degrees of the estimated transformation, and the time in seconds used for key point detection and registration (KDR). The data files used encode the rotation angles θ of the transformations in the filenames, giving ground truth for the evaluation of the performance of our proposed RKP algorithm.

In pattern recognition and information retrieval [29], precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, while recall (also known as sensitivity) is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance. For the keypoint detection in the context of registration of overlapping data and reference shapes, the keypoint detector should repeatably detect any keypoints in the data shape that have also been detected in the reference shape, taking into account that only some will be present due to change of view and occlusion. The overlap of reference and data shapes is defined in terms of their reciprocal correspondences. Because RCs represent correct correspondences and thus the same points on the object of interest, they characterize the repeatability of the detected keypoints in the data and reference shapes. In this case, the precision (Pre) and recall (Rec) rates of the detected keypoints are defined as: $N/n_1 \times 100\%$ and $N/n_2 \times 100\%$, where N, n_1 and n_2 are the number of RCs and the numbers of detected keypoints in the data and reference shapes respectively. Our definition of recall rate is similar to that in [1].

Such definitions of precision and recall rates can be understood from two points of view: (i) the keypoints in the data and reference shapes can be detected and compared individually with the ground truth, showing the extent to which the detected keypoints are relevant to the ground truth. In this case, the keypoints detected in the data and reference shapes are equally treated and play the same role in retrieving the ground truth. Both Pre and Rec rates essentially measure the precision of detected keypoints in the data and reference shapes respectively; (ii) the keypoints to be detected in the data shape should agree with those detected in the reference shape. The precision rate shows the extent to which the detected keypoints in the data shape are repeatable and relevant to those in the reference shape. The recall rate shows the extent to which the keypoints in the reference shape are successfully detected and retrieved by those in the data shape. The distinction between precision and recall rates lies in that they distinguish different roles played by the keypoints in the data and reference shapes: while the former attempts to retrieve and maximize its overlap with the latter, the latter is regarded as ground truth. It can be seen that the second interpretation reflects more realistically the nature of keypoint detection and registration, since no independent ground truth is available for performance measurement. The data and reference shapes are closely tied to the *differences* in representation of the geometry of the object of interest from different viewpoints, and the size of their overlap is defined through registration of the keypoints in the data shape with those in the reference shape. Thus, we adopted this approach throughout this paper.

The experimental study was carried out to evaluate six aspects of the proposed RKP algorithm, as detailed below: key point detection, reference registration, point selection from the data shape and both

shapes, window size and the maximum number of iterations for adaptive smoothing. To facilitate visualization, the estimated transformation was applied to the whole data shape, rather than the keypoints only. Experimental results are presented in Figures 3–10, and Tables 1–7. In Figures 4–8 and Figure 10, yellow represents the transformed data shape, while green represents the reference shape. All the experiments were carried out on a Pentium IV, 2.8GHz computer with 504MB RAM with unoptimized code in Microsoft Visual C++ 6.0.

3.1 Key point detection

In this section, we use real data to demonstrate the detected key points. To this end, the valve20, valve10, valve0, dinosaur72, dinosaur36, dinosaur0, cow49, and cow45 shapes in Figure 1 were selected; results are presented in Figure 2. The key points detected are represented by red plus signs.

While the points detected by the proposed RKP algorithm are distributed relatively evenly over the whole shapes, they do characterize the main features and details of the valve, dinosaur and cow shapes. In particular, more key points were detected on the ports of the valve, the head, belly, toes and tail of the dinosaur, and the ears, eyes and mouth of the cow. Even though the number of key points varies from one shape to another, around 10% points in the original shape are usually detected as key points. This means that around 10% points can be used for a faithful representation of the geometry and details of each 3D freeform shape of interest.

In contrast, both the octave and the MSFE methods selected points mainly in areas with depth discontinuities, since points in these areas usually vary significantly in normal vector and eigenvalues of the local covariance matrix. The NSS method sampled points from both depth discontinuous and planar areas, although more points were selected from depth discontinuous areas with larger changes in normal vector. While it is usually difficult to distinguish foreground objects of interest from the cluttered background without prior knowledge, they were treated equally, and thus key points were selected from both of them. The RKP and NSS methods provide visually more accurate representation of the overall geometry and details for the original valve, dinosaur and cow shapes.

While it typically took under 2s for the RKP, MSFE and NSS methods to detect key points on a shape, the octave method took up to 30s, since it was designed to operate over point clouds, rather than structured range images. This observation is consistent with their computational complexity: while the RKP, MSFE, and NSS methods have similar linear computational complexity in the number of points in the shape, the octave method has quadratic computational complexity as it depends on computation of interpoint distances to find points within a threshold distance of a point of interest.

3.2 Reference registration

We next use the complete sets of points in the original shapes for registration to provide a reference, allowing evaluation of the extent to which the selected key points can accelerate registration and affect its accuracy. To this end, the valve20-10, valve10-0, dinosaur72-36, dinosaur36-0, bottle0-36, bunny80-60, cow49-45, and tubby120-80 shape pairs in Figure 1 were selected. The experimental results are presented in Figure 3 and Table 1. The valve, dinosaur, bunny, and tubby shapes were accurately registered by both the SoftICP and FICP algorithms. The estimated rotation angles for the underlying transformations are close to the ground truth. The transformed data shapes fit onto the reference shapes perfectly. The bottle0-36 and cow59-45 shapes are challenging to register since the former includes a simple cylindrical shape with ambiguity in constraining the underlying transformation, while the latter has a cluttered background, complicating evaluation of the quality of tentative correspondences

established. Nevertheless, such shapes are useful to reveal the true performance of different key point detection techniques. The SoftICP algorithm is more accurate than the FICP algorithm, since the latter registered the dominant bottle body, but not the bottle handle—the probabilistic SoftICP algorithm is more powerful than the threshold based FICP algorithm.

3.3 Point selection from the data shape

We next investigate different point selection techniques applied to the data shapes **P** only, as was done in [7]. In this case, registration was performed between the sampled points in the data shape and *all* points in the reference shape; in this case only the precision rate of the detected key points could be defined, but not the recall rate. The experimental results are presented in Figures 4 and 5 and Tables 2 and 3. It can be seen that the proposed RKP algorithm always detected better key points, with superior precision to the octave, MSFE, and NSS methods of as much as 43%, 36%, and 34% respectively. This higher precision for the RKP key points carries through to more accurate registration results. The SoftICP algorithm accurately registered points selected from all the 8 data shapes: the key points selected by our RKP approach are a useful basis for registration. In contrast, using key points generated by the other methods, the SoftICP algorithm inaccurately registered the bottle0, valve20, dinosaur36, and bunny80 shapes, and failed to register the points selected by either the octave or MSFE method for the cow49 shape and the points selected by the NSS method for the tubby120 shape. For bottle0, the SoftICP algorithm registered the dominant bottle body, rather than the bottle handle. Over all 8 pairs of overlapping shapes, the increase in average registration error compared to using full data without point sampling was 6%, 31%, 29%, and 16% for the RKP, octave, MSFE, and NSS methods respectively. The octave, MSFE, and NSS methods have reduced precision compared to the proposed RKP algorithm for detection of key points by up to 26%, 45%, and 30% respectively: the key points detected by the former are less repeatably placed and thus less useful as a basis for registration. While the FICP algorithm failed to register the points selected by all four methods for the bottle0 shape, it successfully registered the points selected by our proposed RKP algorithm for all 7 other shapes. In contrast, it failed to register the points selected by the octave, MSFE, and NSS methods for the dinosaur72, dinosaur36, and cow49 shapes, and points selected by the MSFE method for the data tubby120 shape. The registration failures can be seen in the legs of the dinosaur and the head of the cow in shapes dinosaur72, dinosaur36, and cow49 relative to the references dinosaur36, dinosaur0, and cow45 respectively. Over all 8 pairs of overlapping shapes, the average registration error was increased compared to using full data without point sampling by 6%, 60%, 116%, and 71% by the RKP, octave, MSFE, and NSS methods respectively.

The above analysis shows that sampling around 10% points using the RKP algorithm has little effect on the registration accuracy for overlapping 3D freeform shapes, especially when using the FICP algorithm. The octave, MSFE, and NSS methods do not perform as well, as they have to estimate quantities based on second-order derivatives of the discrete range data, such as normal vectors and eigenvalues of the local covariance matrix, and these are sensitive to imaging noise, occlusion, and appearance and disappearance of points. In contrast, the RKP algorithm employs adaptive smoothing and first order derivatives, and thus is more robust. The adaptive smoothing operation provides a reliable reference for the RKP algorithm to select key points as local maxima.

Comparing Tables 2 and 3 with Table 1, it can be seen that the SoftICP algorithm is more accurate than the FICP algorithm. Both algorithms produce worse results when using point sampling, as might be expected. By comparison, the FICP algorithm is more sensitive to use of point sampling than the SoftICP algorithm. This is because the entropy maximisation principle, the two-way constraint, and the deterministic annealing scheme in the SoftICP algorithm provide a powerful probabilistic framework for

Table 2: The precision Pre rate of the detected keypoints, the average e_μ and standard deviation e_σ of registration errors in millimetres based on RCs, expected and estimated rotation angles θ and $\hat{\theta}$ in degrees, and KDR time t in seconds for points selected from the data shape using different algorithms and registered using the SoftICP algorithm.

Image	Algo.	Pre (%)	e_μ (mm)	e_σ (mm)	θ (°)	$\hat{\theta}$ (°)	t (s)
valve20-10	RKP	83.85	0.42	0.23	10	10.08	9
	octave	75.90	0.43	0.22		9.94	11
	MSFE	64.10	0.50	0.24		9.71	13
	NSS	63.32	0.46	0.23		9.87	17
valve10-0	RKP	94.06	0.39	0.20	10	10.14	5
	octave	89.44	0.40	0.21		10.09	16
	MSFE	86.46	0.39	0.20		10.04	5
	NSS	81.30	0.40	0.20		10.07	4
dinosaur72-36	RKP	74.08	0.64	0.85	36	35.32	4
	octave	64.93	0.76	0.84		34.45	4
	MSFE	55.63	0.74	0.88		34.08	6
	NSS	50.14	0.64	0.85		34.86	7
dinosaur36-0	RKP	78.30	0.59	0.54	36	35.45	8
	octave	63.87	0.60	0.55		35.58	7
	MSFE	55.38	0.72	0.58		35.22	19
	NSS	51.55	0.68	0.56		34.83	9
bottle0-36	RKP	80.08	0.78	0.47	36	29.84	3
	octave	71.75	0.68	0.34		32.70	4
	MSFE	63.51	0.67	0.33		33.32	5
	NSS	78.07	0.68	0.36		32.63	3
bunny80-60	RKP	90.88	0.23	0.11	20	20.07	4
	octave	71.36	0.27	0.12		19.22	4
	MSFE	78.02	0.24	0.11		19.51	4
	NSS	65.32	0.28	0.14		18.76	4
cow49-45	RKP	70.25	0.74	0.70	40	40.92	3
	octave	24.76	1.57	2.61		91.39	5
	MSFE	22.22	1.31	2.85		62.61	4
	NSS	42.54	0.96	0.98		41.77	4
tubby120-80	RKP	79.07	0.26	0.18	40	38.86	4
	octave	61.78	0.28	0.16		38.64	15
	MSFE	52.62	0.38	0.23		38.02	4
	NSS	52.62	0.34	0.22		29.04	3

Table 3: The precision Pre rate of the detected keypoints, the average e_μ and standard deviation e_σ of registration errors in millimetres based on RCs, expected and estimated rotation angles θ and $\hat{\theta}$ in degrees, and KDR time t in seconds for points selected from the data shape using different algorithms and registered using the FICP algorithm.

Image	Algo.	Pre (%)	e_μ (mm)	e_σ (mm)	θ (°)	$\hat{\theta}$ (°)	t (s)
valve20-10	RKP	83.63	0.42	0.23	10	10.10	5
	octave	76.11	0.42	0.22		10.07	5
	MSFE	63.32	0.45	0.23		9.89	7
	NSS	63.11	0.44	0.23		9.93	17
valve10-0	RKP	93.61	0.39	0.21	10	10.16	4
	octave	89.51	0.39	0.21		10.07	4
	MSFE	85.68	0.39	0.20		10.07	3
	NSS	83.72	0.45	0.20		9.83	3
dinosaur72-36	RKP	75.31	0.71	0.84	36	34.14	3
	octave	46.33	1.19	1.01		22.11	2
	MSFE	34.51	1.17	0.78		16.05	4
	NSS	35.07	1.23	1.02		18.59	4
dinosaur36-0	RKP	78.30	0.60	0.55	36	35.35	5
	octave	41.50	1.44	1.27		11.42	7
	MSFE	28.47	1.30	1.20		9.38	5
	NSS	39.47	1.46	1.29		14.74	5
bottle0-36	RKP	80.49	0.70	0.47	36	9.88	4
	octave	73.51	0.75	0.69		21.28	3
	MSFE	26.14	2.78	3.30		36.01	5
	NSS	77.89	0.80	0.66		25.85	4
bunny80-60	RKP	90.37	0.23	0.11	20	20.11	4
	octave	70.59	0.28	0.13		19.43	3
	MSFE	77.71	0.25	0.11		19.17	3
	NSS	66.25	0.32	0.17		18.30	3
cow49-45	RKP	68.46	0.74	0.70	40	40.73	4
	octave	23.81	1.30	1.71		76.52	4
	MSFE	12.38	1.29	1.63		30.67	3
	NSS	26.34	1.44	1.41		31.64	3
tubby120-80	RKP	78.49	0.27	0.18	40	38.83	3
	octave	58.38	0.38	0.17		37.94	3
	MSFE	28.27	0.68	0.59		41.15	3
	NSS	55.76	0.43	0.28		25.50	3

Figure 4: Registration results using the SoftICP algorithm for various shapes with key points selected from the data shape using different point selection algorithms. Rows, top to bottom: RKP, octave, MSFE, and NSS. Columns, left to right: valve20-10, valve10-0, dinosaur72-36, dinosaur36-0, bottle0-36, bunny80-60, cow49-45, and tubby120-80.

estimation for the weights of the tentative correspondences established, while the FICP algorithm has difficulty in defining the quality of tentative correspondences and thus in rejecting outliers. The FICP

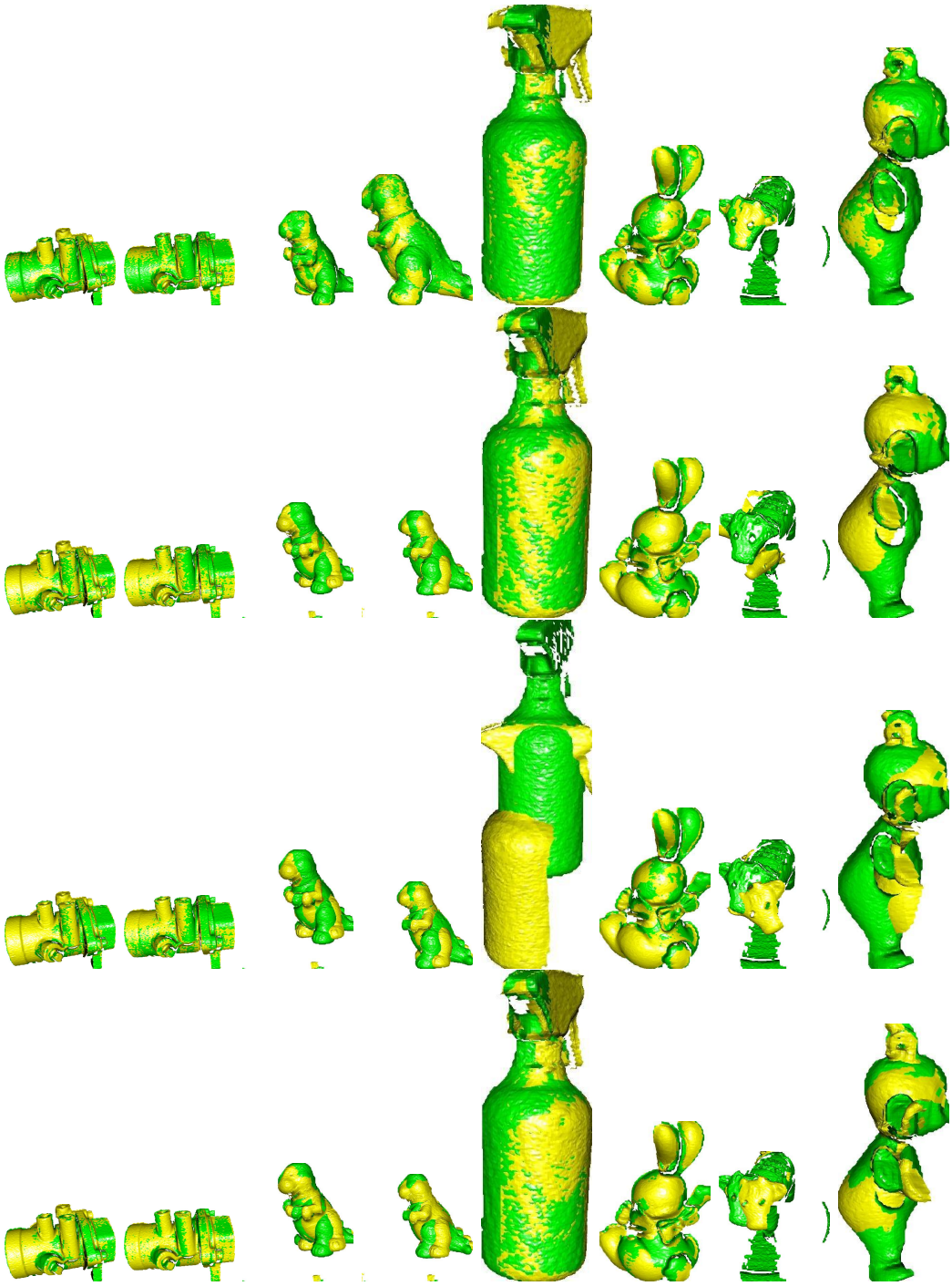

Figure 5: Registration results using the FICP algorithm for various shapes with key points selected from the data shape using different point selection algorithms. Rows, top to bottom: RKP, octave, MSFE, and NSS. Columns, left to right: valve20-10, valve10-0, dinosaur72-36, dinosaur36-0, bottle0-36, bunny80-60, cow49-45, and tubby120-80.

algorithm is slightly more efficient than the SoftICP algorithm, but it usually converges prematurely, producing inaccurate registration results. While all registration results are slightly worse when sampling,

the time needed for registration is reduced by as much as 78% using point sampling and registration by the RKP algorithm. Detecting key points speeds up registration without significant loss of accuracy.

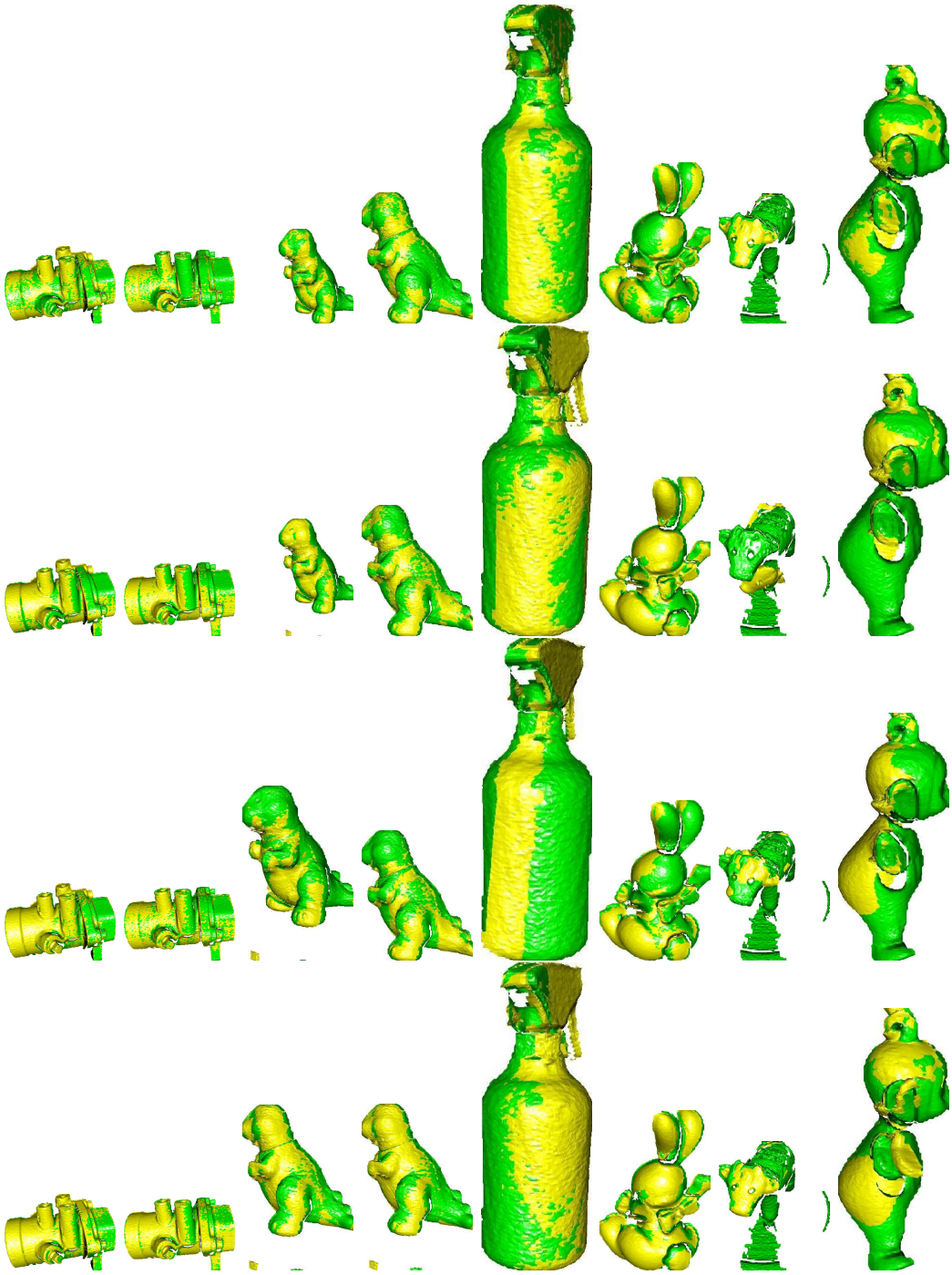

Figure 6: Registration results using the SoftICP algorithm for various shapes using key points selected from *both* shapes by different algorithms. Rows, top to bottom: RKP, octave, MSFE, and NSS. Columns, left to right: valve20-10, valve10-0, dinosaur72-36, dinosaur36-0, bottle0-36, bunny80-60, cow49-45, and tubby120-80.

Table 4: Precision Pre and recall Rec rates of the detected keypoints, the average e_μ and standard deviation e_σ of registration errors in millimetres based on RCs, expected and estimated rotation angles θ and $\hat{\theta}$ in degrees, and KDR time t in seconds, for points selected from both shapes using various algorithms and registered using the SoftICP algorithm.

Image	Algo.	Pre (%)	Rec (%)	e_μ (mm)	e_σ (mm)	θ $(^\circ)$	$\hat{\theta}$ $(^\circ)$	t (s)
valve20-10	RKP	48.45	59.13	0.41	0.22	10	10.08	5
	octave	35.62	39.43	0.44	0.23		10.08	6
	MSFE	45.86	50.78	0.47	0.23		9.86	6
	NSS	40.71	45.07	0.47	0.23		9.91	6
valve10-0	RKP	59.04	55.08	0.40	0.21	10	10.01	4
	octave	42.80	43.07	0.41	0.20		10.09	5
	MSFE	57.51	57.87	0.40	0.20		10.04	7
	NSS	53.52	53.85	0.39	0.20		10.07	7
dinosaur72-36	RKP	39.88	35.71	0.65	0.86	36	35.04	6
	octave	25.63	21.77	0.67	0.86		34.95	6
	MSFE	34.65	29.42	0.73	0.84		34.75	4
	NSS	30.56	25.96	1.11	1.01		31.21	4
dinosaur36-0	RKP	42.58	39.14	0.59	0.54	36	35.45	8
	octave	25.71	24.77	0.61	0.55		35.33	4
	MSFE	29.07	27.99	0.69	0.57		36.00	4
	NSS	31.94	30.76	0.89	0.65		33.68	4
bottle0-36	RKP	42.91	43.27	0.75	0.32	36	34.80	3
	octave	26.31	27.62	0.82	0.72		20.94	3
	MSFE	27.54	28.91	1.85	1.42		32.43	4
	NSS	25.44	26.70	0.88	0.65		26.13	3
bunny80-60	RKP	53.21	50.16	0.23	0.11	20	19.91	4
	octave	25.54	23.91	0.31	0.13		19.53	3
	MSFE	51.08	47.83	0.26	0.11		19.52	3
	NSS	41.02	38.40	0.31	0.14		18.86	3
cow49-45	RKP	39.43	15.99	0.79	0.88	40	40.38	3
	octave	9.21	4.09	1.93	3.00		79.78	3
	MSFE	21.90	9.73	0.91	1.28		51.83	2
	NSS	7.94	3.53	1.29	2.41		104.81	3
tubby120-80	RKP	34.59	28.54	0.28	0.18	40	39.06	2
	octave	22.77	20.00	0.40	0.21		39.70	2
	MSFE	32.98	28.96	0.45	0.30		37.49	2
	NSS	31.94	28.05	0.47	0.34		29.43	4

Table 5: Precision Pre and recall Rec rates of the detected keypoints, the average e_μ and standard deviation e_σ of registration errors in millimetres based on RCs, expected and estimated rotation angles θ and $\hat{\theta}$ in degrees, and KDR time t in seconds for points selected from both shapes using different algorithms, and registered using the FICP algorithm.

Image	Algo.	Pre (%)	Rec (%)	e_μ (mm)	e_σ (mm)	θ $(^\circ)$	$\hat{\theta}$ $(^\circ)$	t (s)
valve20-10	RKP	48.67	59.40	0.42	0.23	10	10.02	4
	octave	36.11	39.98	0.45	0.23		10.12	4
	MSFE	45.94	50.86	0.44	0.22		10.04	4
	NSS	38.43	43.66	0.55	0.28		10.03	5
valve10-0	RKP	58.95	54.99	0.40	0.20	10	9.97	3
	octave	42.95	43.22	0.43	0.20		10.12	4
	MSFE	57.35	57.71	0.39	0.21		10.07	4
	NSS	52.82	53.15	0.41	0.21		10.07	5
dinosaur72-36	RKP	39.26	35.16	0.66	0.87	36	35.20	4
	octave	18.87	16.03	1.19	0.91		22.24	2
	MSFE	16.90	14.35	1.19	1.01		22.70	3
	NSS	19.44	16.51	1.23	0.97		19.87	3
dinosaur36-0	RKP	42.44	39.01	0.60	0.55	36	35.35	6
	octave	15.55	14.97	1.51	1.31		8.72	4
	MSFE	11.12	10.71	1.48	1.13		8.53	4
	NSS	22.49	21.65	1.33	1.26		14.09	4
bottle0-36	RKP	44.15	44.51	0.87	0.58	36	19.72	3
	octave	24.21	25.41	1.09	0.61		23.36	3
	MSFE	24.21	25.41	2.11	1.62		27.42	2
	NSS	23.50	24.68	1.44	0.80		20.19	3
bunny80-60	RKP	53.21	50.16	0.27	0.13	20	18.82	2
	octave	25.70	24.06	0.31	0.14		19.34	2
	MSFE	52.01	48.69	0.25	0.12		18.76	3
	NSS	39.16	36.67	0.36	0.19		18.30	3
cow49-45	RKP	39.78	16.13	0.85	0.85	40	44.01	3
	octave	6.98	3.10	1.72	1.20		75.27	3
	MSFE	22.86	10.15	0.78	0.90		45.01	3
	NSS	11.43	5.08	1.60	1.41		34.86	3
tubby120-80	RKP	34.01	28.06	0.29	0.18	40	38.45	3
	octave	21.20	18.62	0.43	0.22		40.04	3
	MSFE	19.63	17.24	0.64	0.42		30.58	3
	NSS	30.37	26.67	0.67	0.37		29.67	4

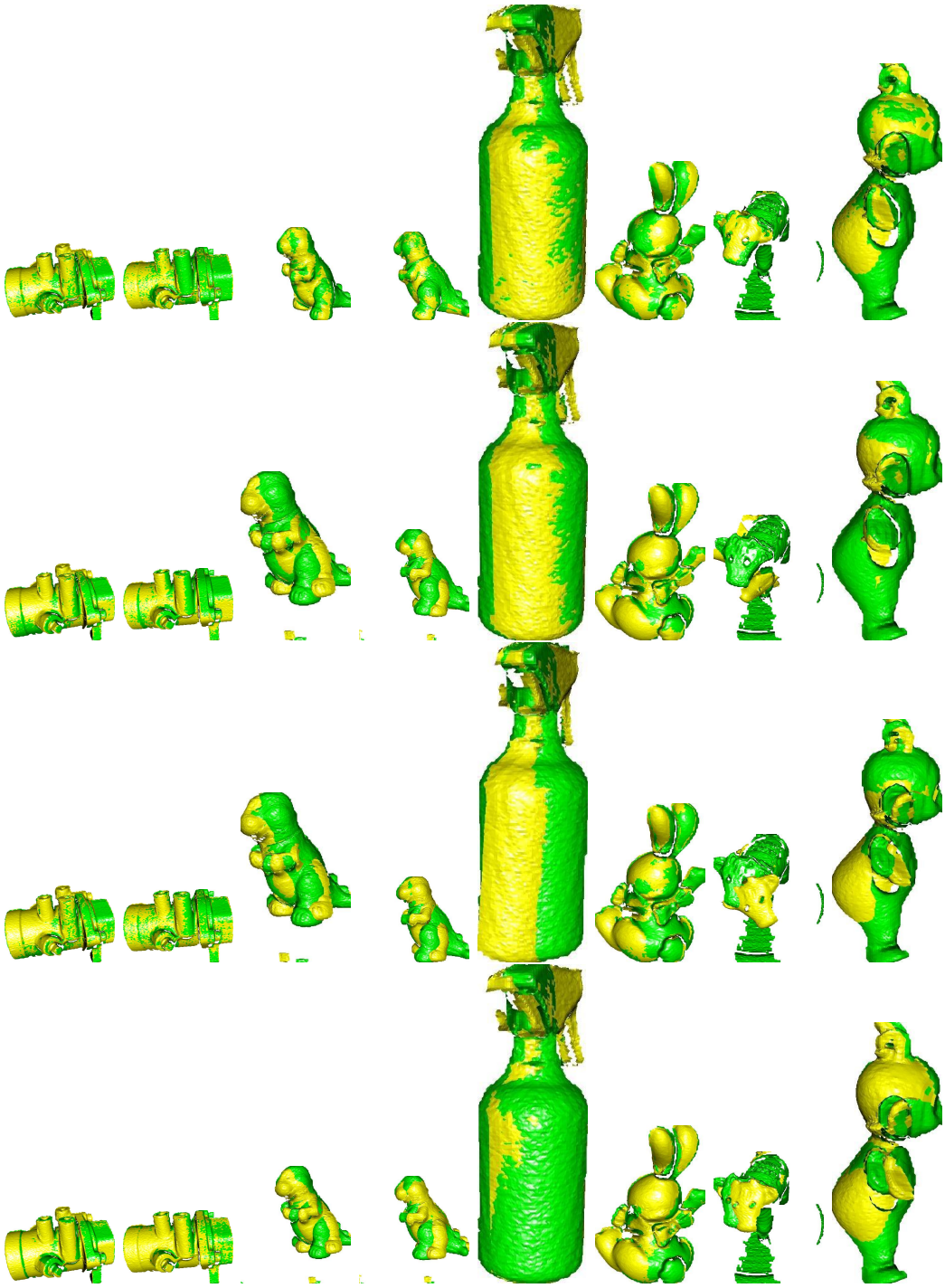

Figure 7: Registration results using the FICP algorithm for various shapes using key points selected from *both* shapes by different algorithms. Rows, top to bottom: RKP, octave, MSFE, and NSS. Columns, left to right: valve20-10, valve10-0, dinosaur72-36, dinosaur36-0, bottle0-36, bunny80-60, cow49-45, and tubby120-80.

3.4 Point selection from both shapes

In the previous experiments, only the data shapes were sampled with key points. Here, we sample both the data and reference shapes and again perform registration using both the SoftICP and FICP algo-

rithms. Such sampled points are more challenging to register, since they are less likely to correspond to exactly the same points on the shapes, and failures to robustly identify the same key points will become apparent. To produce accurate registration results, the sampled points should faithfully represent the geometry and details of the shapes of interest from any viewpoint, and should reliably represent the original shapes from different viewpoints. Our experimental results are presented in Figures 6 and 7 and Tables 4 and 5. It can be seen that the precision rate of the RKP algorithm is always higher than for other algorithms, and its recall rate is also higher except in two cases, for the valve10-0 and tubby120-80 shape pairs, where they are similar to those produced by the MSFE method. Over all 8 pairs of overlapping shapes, the proposed RKP algorithm had better precision and recall rates compared to the octave, MSFE, and NSS methods on average by as much as 68%, 55%, 19%, 13%, 36%, and 26% respectively, showing that the key points detected by the RKP method are more representative of the underlying geometry. The higher precision and recall rates of these key points are confirmed by the registration results. The SoftICP algorithm accurately registered the points sampled by the proposed RKP method with the average error increasing by as little as 7%, even though both the data and reference shapes were sampled, reducing time by up to 90%. In contrast, point sampling by the octave, MSFE, and NSS methods increased errors by up to 46%, 51%, and 52% respectively. Even when sampling both shapes by the RKP algorithm, good registration can be achieved, and all the overlapping shapes were brought into accurate alignment with each other. In contrast, the sampled points from the octave, MSFE, and NSS methods prove less useful for registration; the transformed valve20, dinosaur72 and tubby120 shapes are displaced in 3D space with respect to the valve10, dinosaur36 and tubby80 shapes respectively.

Again, when performing registration with the FICP algorithm, the RKP algorithm always produced the highest precision and recall rates of any key point selection method, except in one case, when a lower recall rate was obtained for the valve10-0 pair than with the MSFE method. The RKP algorithm is better than the octave, MSFE, and NSS methods for the detection of key points in the sense of precision and recall rates by as much as 88%, 76%, 44%, 39%, 51%, and 43% respectively. These remarkable results show that out of these methods, the RKP method detected key points most closely representing the geometry and details of the original shapes, providing the best registration results. Even though point sampling by the octave, MSFE, and NSS methods increased the average registration error significantly, by as much as 85%, 89%, and 97% respectively, for the RKP method it only increased by just 13%. While the FICP algorithm failed to register the points sampled by the octave, MSFE, and NSS methods, causing the transformed dinosaur72, dinosaur36, cow49, and tubby120 shapes to intersect in 3D space the dinosaur36, dinosaur0, cow45, and tubby80 shapes respectively, it successfully registered the points sampled by the RKP algorithm and brought all the overlapping shapes into accurate alignment, except for the bottle0-36 pair, which was more challenging to register due to their simple shapes.

The SoftICP algorithm exhibits similar behaviors for the registration of the points sampled from both the data and reference shapes by the RKP, octave, and MSFE methods to those sampled from the data shape only, but it produces worse results when registering points sampled by the NSS method. The repeatability of points sampled by the NSS method is worse due to the random nature of its sampling. The FICP algorithm is more sensitive to the point sampling.

3.5 Window size

In other experiments with the RKP algorithm, we set the window size to 3×3 for neighbor detection and adaptive smoothing. In this section, we experimentally investigate whether this is a good option and consider window sizes from 3×3 to 7×7, using the new duck0-20, frog0-20, lobster0-20, and buddha0-

Table 6: Precision Pre and recall Rec rates of the detected keypoints, the average e_μ and standard deviation e_σ of registration errors in millimetres based on RCs, expected and estimated rotation angles θ and $\hat{\theta}$ in degrees, and KDR time t in seconds for points selected from both shapes by the proposed RKP algorithm using different window sizes and registered by different algorithms.

Image	Size	Algo.	Pre (%)	Rec (%)	e_μ (mm)	e_σ (mm)	θ (°)	$\hat{\theta}$ (°)	t (s)
duck0-20	3×3	SoftICP	52.00	52.20	0.41	0.29	20	18.99	4
		FICP	51.11	51.30	0.41	0.26		18.69	5
	7×7	SoftICP	44.34	44.51	0.44	0.27		18.33	5
		FICP	45.08	45.24	0.47	0.30		16.39	5
frog0-20	3×3	SoftICP	40.71	49.15	0.31	0.16	20	19.40	4
		FICP	41.14	49.67	0.32	0.17		18.71	3
	7×7	SoftICP	39.87	45.82	0.32	0.16		19.14	4
		FICP	40.65	46.72	0.34	0.18		18.47	3
lobster0-20	3×3	SoftICP	50.80	54.40	0.38	0.23	20	19.17	4
		FICP	50.80	54.40	0.41	0.25		18.44	3
	7×7	SoftICP	47.34	47.69	0.38	0.24		19.22	4
		FICP	47.20	47.54	0.40	0.26		18.50	3
buddha0-20	3×3	SoftICP	56.17	54.10	0.60	0.24	20	19.88	6
		FICP	50.57	48.70	0.94	0.57		8.07	4
	7×7	SoftICP	53.68	48.88	0.87	0.50		13.33	16
		FICP	42.83	39.00	0.88	0.50		13.21	6

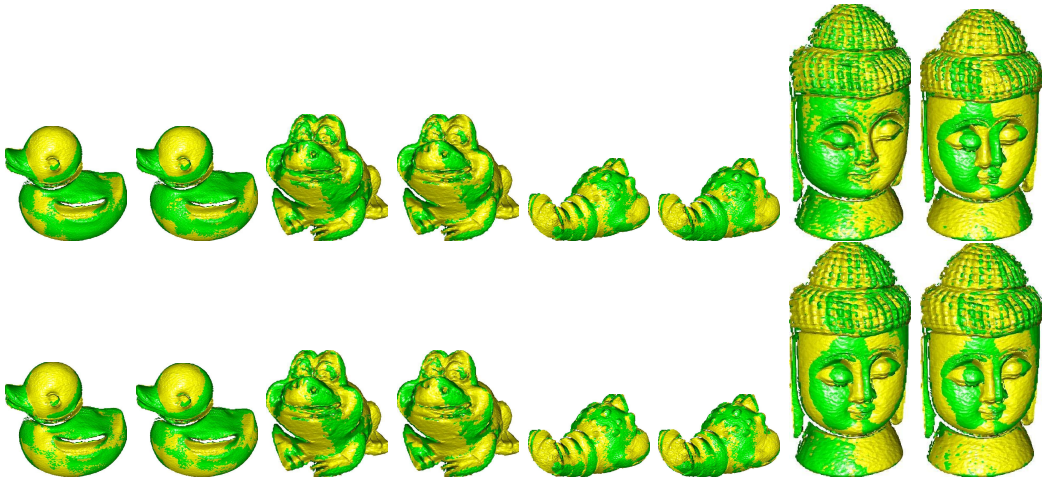

Figure 8: Registration results of different algorithms with the keypoints selected from both shapes using the RKP algorithm with the window size taking different values. Rows, top to bottom: 3×3, 7×7 windows. Models, left to right: duck0-20, frog0-20, lobster0-20, and buddha0-20. Odd columns: SoftICP; even columns: FICP.

20 shape pairs illustrated in Figure 1. Both data and reference shapes were sampled and registered using both the SoftICP and FICP algorithms. The experimental results are presented in Figure 8 and Table 6. A larger window size usually decreases the precision and recall rates of the detected key points registered by the SoftICP algorithm on average by 7% and 10%, and by the FICP algorithm by 9% and 12% respectively. A larger window size results in less reliable key points, because it blurs local features and thus makes key point detection harder. This conclusion is confirmed by the registration results. A larger window size of 7×7 usually produces slightly worse results than when using 3×3; the rotation angle of the estimated transformation is close to the ground truth. Significantly worse results are obtained when registering the boddha0-20 pair using the SoftICP algorithm: the nose, eyes, and ears in the transformed data buddha0 and reference buddha20 shapes are clearly displaced in 3D space with the average error being increased by as much as 45%. Thus, a window size of 3×3 is suggested for key point detection.

3.6 The maximum number of iterations for adaptive smoothing

In this section, we investigate the determination of the maximum number ($I_{\max}$) of iterations used for adaptive smoothing in our RKP algorithm. To this end, we considered three possibilities: 10, 30, and 50. The new free form shapes angel0, angle20, angle40, bird0, and bird20 illustrated in Figure 1 were selected for the experiments, with point sampling applied to both shapes which were registered using both the SoftICP and FICP algorithms again. The experimental results are presented in Figures 9 and 10 and Table 7.

Figure 9 and Table 7 show that the larger the maximum number of iterations for adaptive smoothing, the fewer keypoints the proposed RKP algorithm detects. This is because when the maximum number of iterations is small, 10, for example, the smoothing operation takes place mainly locally, leading points to be compared in a small area and thus more points to be depth maxima and to be selected as keypoints. In contrast, when the number is large, 50, for example, the smoothing operation propagates local information from one region to another, enabling comparison of points over a larger area, leading

Table 7: Precision Pre and recall Rec rates of the detected keypoints, the average e_μ and standard deviation e_σ of registration errors in millimetres based on RCs, expected and estimated rotation angles θ and $\hat{\theta}$ in degrees, and KDR time t in seconds for keypoints selected from both shapes by the proposed RKP algorithm with the parameter $I_{\max}$ taking different values and registered by different algorithms. n_1 and n_2 are the numbers of the detected points in the data and reference shapes respectively.

Image	$I_{\max}$ (n_1, n_2)	Algo.	Pre (%)	Rec (%)	e_μ (mm)	e_σ (mm)	θ (°)	$\hat{\theta}$ (°)	t (s)
angel0-20	10	SoftICP	54.29	50.37	0.49	0.26	20	20.52	7
	(2365, 2549)	FICP	54.04	50.14	0.50	0.25		20.48	3
	30	SoftICP	50.34	45.84	0.50	0.26		20.48	4
	(1301, 1429)	FICP	50.65	46.12	0.52	0.26		20.34	2
	50	SoftICP	49.95	44.62	0.50	0.26		20.54	4
	(989, 1107)	FICP	50.05	44.71	0.51	0.27		20.42	2
angel20-40	10	SoftICP	46.72	54.99	0.54	0.34	20	20.35	6
	(2549, 2166)	FICP	46.92	55.22	0.55	0.34		20.22	4
	30	SoftICP	45.55	53.14	0.55	0.34		20.35	4
	(1429, 1223)	FICP	45.91	53.55	0.57	0.34		20.21	2
	50	SoftICP	43.00	50.64	0.55	0.34		20.34	3
	(1107, 940)	FICP	43.54	51.28	0.59	0.34		20.10	3
bird0-20	10	SoftICP	43.17	56.13	0.29	0.12	20	19.56	4
	(1538, 1183)	FICP	42.52	55.28	0.34	0.13		19.04	4
	30	SoftICP	39.75	53.51	0.30	0.13		19.67	3
	(883, 656)	FICP	40.09	53.96	0.39	0.15		18.15	3
	50	SoftICP	38.23	50.29	0.32	0.13		19.79	3
	(667, 507)	FICP	39.73	52.27	0.45	0.19		17.68	3

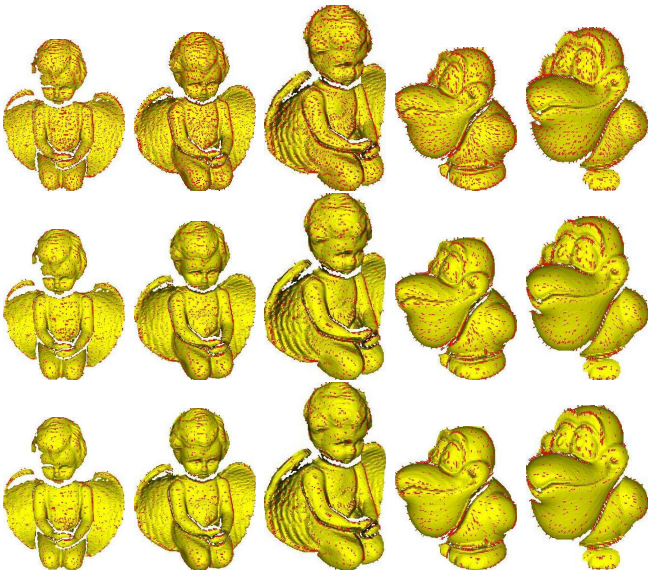

Figure 9: The keypoints detected by the proposed RKP algorithm with the parameter $I_{\max}$ taking different values. Rows, top to bottom: $I_{\max}$=10, 30, and 50. Models, left to right: angel0, angel20, angel40, bird0 and bird20.

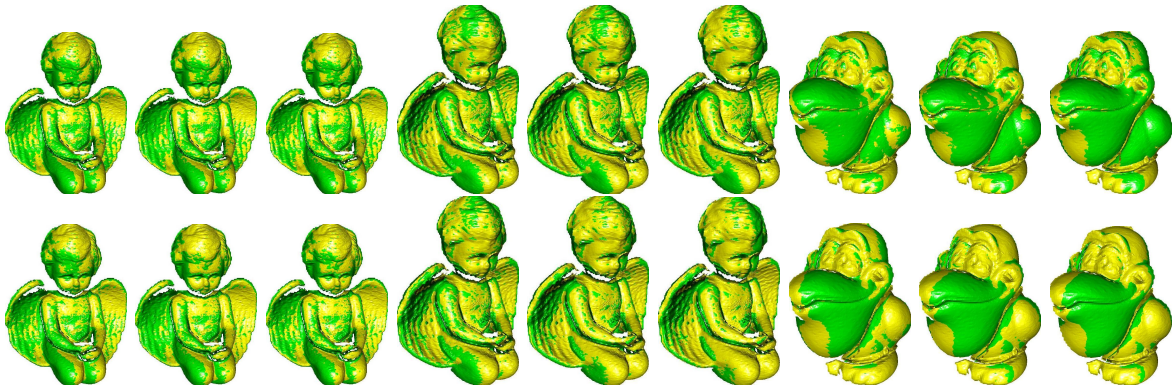

Figure 10: Registration results of different algorithms with the keypoints selected from both shapes using the proposed RKP algorithm with the parameter $I_{\max}$ taking different values. Rows, top to bottom: SoftICP and FICP. Models, left three: angel0-20; Middle three: angel20-40; Right three: bird0-20. Columns 1, 4, and 7: $I_{\max}$=10; Columns 2, 5, and 8: $I_{\max}$=30; Columns 3, 6, and 9: $I_{\max}$=50.

to fewer points being selected as keypoints with maximal depths. Figure 10 and Table 7 show that while a large number of points takes more time to process and register, they tend to produce more accurate registration results, with increased precision and recall rates for the detected keypoints: they describe the geometry and details of the underlying shapes more faithfully. In contrast, a smaller number of points is more computationally efficient to register, but usually produces larger registration errors, as fewer points less well characterize the geometry of the free form shapes. This observation has been demonstrated by the fact that the FICP algorithm superimposes the transformed bird0 shape over the bird20 shape with less inter-penetration through each other, and increases the average error of that when $I_{\max} = 10$ by as much as 32%; the number of the detected keypoints drops rapidly by 58% and 59% in the angel0 and angel20 shapes with the increase of the maximum number of iterations for adaptive

smoothing from 10 to 50, while the drop in the precision and recall rates is less serious, just 8% and 11% registered by the SoftICP algorithm with a gain of 42% in overall time. These results show that the proposed RKP method: (i) can flexibly detect varying numbers of keyptoints as required by controlling the value of $I_{\max}$, and (ii) can stably detect repeatable key points. $I_{\max} = 30$ is recommended as a good compromise between registration accuracy and computational efficiency.

4 Conclusions

While the latest laser scanners enable fast and affordable capture of depth maps and shapes of interest, they must be registered to form complete models. Registration is a costly process, and in this paper, we show to significantly reduce this cost using a sampling method based on a novel way of selecting key points. Our contributions can be summarized as follows.

- Inspired by retinex theory [6, 16], we have proposed a novel method of key point detection using adaptive smoothing. This operation suppresses the impact of points in depth discontinuous regions, and all points unaffected by smoothing are selected as key points. Key points have locally maximal transformed and normalized depth. The method is easy to implement. To the best of our knowledge, this is the first time that retinex theory has been adapted from 2D image enhancement for application in 3D key point analysis and detection.

- The detected key points have been shown to be useful for efficient registration of overlapping 3D freeform shapes. No matter whether key points are selected from just one or both shapes, accurate results have been obtained, with registration being up to 20 times faster than when using unsampled shapes. This is a significant improvement in computational efficiency without sacrificing registration accuracy. The reason why our proposed RKP algorithm works well in this context is that the smoothed depths provide a reliable reference for the judgment of whether a point is a key point. Existing methods detect salient points in an ad hoc way, while the proposed RKP method guarantees that the detected points are locally highest ones after normalization and adaptive smoothing.

- A comparative study has been performed between the proposed key point detection method and three other state-of-the-art methods, using real depth images. It shows that it is feasible to apply ICP variants to directly register key points as long as they are informative, expressive, and repeatable enough. The precision and recall rates of the key points detected by the proposed RKP method are better than those for the selected competitors by as much as 88% and 76%. RKP sampling of both shapes increases the average registration error by as little as 7%, while for other selected state-of-the-art methods it increased by as much as 52%.

Our proposed RKP method is a powerful approach to key point detection. The detected key points are useful in the context of registration, and with direct application of ICP variants, they can produce accurate registration results with significantly improved computational efficiency. Future research will investigate the similarity between the shapes defined by these key points and the original complete point sets, how the detected key points can be applied for the generation of levels of detail for efficient data transmission, rendering and visualization, and how the detected points can be incorporated into feature extraction and matching methods [7, 23] for applications such as registration.

Acknowledgments

We would like to express our sincere thanks to the reviewers for their insightful comments that have improved the quality of the paper, and also to support of the Higher Education Funding Council for Wales (HEFCW) for the One Wales Research Institute of Visual Computing (RIVIC).

References

[1] H. Aanas, A.L. Dahl, K.S. Pedersen. Interesting interest points. *IJCV* 97(2012) 18-35.

[2] U. Castellani, M. Cristani, S. Fantoni and V. Murino. Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Computer Graphics Forum* 27(2008) 643-652.

[3] L. Chen, L. Zhang, H. Zhang, M. Abdel-Mottaleb. 3D Shape constraint for facial feature localization using probabilistic-like output. *Proc. the Sixth IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 302-307, 2004.

[4] C. Creusot, N. Pears, J. Austin. Automatic keypoint detection on 3D faces using a dictionary of local shapes. *Proc. 3DIMPVT*, pp. 204-211, 2011.

[5] S. Gold, A. Rangarajan, C.-P. Lu, S. Pappu, E. Mjolsness. New algorithms for 2-D and 3-D point matching: pose estimation and correspondence. *Pattern Recognition* 31(1998) 1019-1031.

[6] D.J. Jobson, Z. Rahman, G.A. Woodwell. A multiscale Retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Processing* 6(1997) 965-976.

[7] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 21(1999) 433-449.

[8] R. Kimmel, M. Elad,, D. Shaked, and R. Keshet A variational framework for Retinex. *International Journal of Computer Vision*, 52(1), 7-23, 2003

[9] Y. Liu. Automatic 3d freeform shape matching using the graduated assignment algorithm. *Pattern Recognition*, 38(2005) 1615-1631.

[10] Y. Liu. Constraints for closest point finding. *Pattern Recognition Letters*, vol. 29, no. 7, pp. 841-851, 2008.

[11] T.R. Lo, and J.P. Siebert. Local feature extraction and matching on range images: 2.5D SIFT. *Computer Vision and Image Understanding*, 113 (2009) 1235-1250.

[12] A. Mandow, J.L. Martínez, A.J. Reina, Jesús Morales. Fast range-independent spherical subsampling of 3D laser scanner points and data reduction performance evaluation for scene registration. *Pattern Recognition Letters* 31(2010) 1239-1250.

[13] A.S. Mina, M. Bennamoun, R. Owens. Keypoint detection and local feature matching for textured 3D face recognition. *IJCV* 79(2008) 1-12.

[14] J. Novatnack and K. Nishino. Scale-dependent/invariant local 3D shape descriptors for fully automatic registration of multiple sets of range images. *Proc. of the 10th European Conference on Computer Vision (ECCV)*, pp. 440-453, 2008.

[15] OSU(MSU/WSU) range image database. http://sampl.ece.ohio-state.edu/data/3DDB/RID/index.htm, 2010.

[16] Y.K. Park, S.L. Park, J.K. Kim. Retinex method based on adaptive smoothing for illumination invariant face recognition. *Signal Processing* 88(2008) 1929-1945.

[17] M. Pauly, R. Keiser, M. Gross. Multi-scale feature extraction on point-sampled surfaces. *Computer Graphics Forum*, vol. 22, no.3, pp. 281-289, 2003.

[18] J.M. Phillips, R. Liu, C. Tomasi. Outlier Robust ICP for Minimizing Fractional RMSD. *Proc. 3DIM*, pp. 427-434, 2007.

[19] C. Plagemann, V. Ganapathi, D. Koller, S. Thrun,. Real-time identification and localization of body parts from depth images. *Proc. ICRA*, pp. 3108-3113, 2010.

[20] S. Rusinkiewicz, M. Levoy. Efficient variants of the ICP algorithm. *Proc. International Conference on 3D Digital Imaging and Modeling (3DIM)*, pp. 145-152, 2001.

[21] C.-T. Shen, W.-L. Hwang. Color image enhancement using retinex with robust envelope. In *Proc. ICIP*, 2009, pp. 3141-3144.

[22] I. Sipiran1 and B. Bustos. A robust 3D key points detector based on Harris operator. *Proc. Eurographics Workshop on 3D Object Retrieval*, pp. 7-14, 2010.

[23] E.R. Smith, R.J. Radke, C.V. Stewwart. Physical scale keypoints: matching and registration for combined intensity/range images. *International Journal of Computer Vision*, 97(2012) 2-17.

[24] B. Steder, R.B. Rusu, K. Konolige, W. Burgard. Point feature extraction on 3D range scans taking into account object boundaries. *Proc. ICRA*, pp. 2601-2608, 2011.

[25] B. Steder, G. Grisetti, and W. Burgard. Robust place recognition for 3D range data based on point features. *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 1400-1405, 2010.

[26] J. Stuckler and S. Behnke. Key point detection in depth images through scale-space surface analysis. *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3568-3574, 2011

[27] R. Toldo, A. Beinat, F. Crosilla. Global registration of multiple point clouds embedding the Generalized Procrustes analysis into an ICP framework. *Proc. 3DPVT*, 2010.

[28] A. Torsello, E. Rodolà, and A. Albarelli. Sampling relevant points for surface registration. *Proc. 3DIMPVT*, pp. 290-295, 2011.

[29] Precision and recall. http://en.wikipedia.org/wiki/Precision_and_recall